Abstract
The availability of entire genome sequences and the wealth of literature on gene regulation have enabled researchers to model an organism’s transcriptional regulation system in the form of a network. In such a network, TFs (transcription factors) and TGs (target genes) are represented as nodes and regulatory interactions between TFs and TGs are represented as directed links. In the present review, I address the following topics pertaining to transcriptional regulatory networks. (i) Structure and organization: first, I introduce the concept of networks and discuss our understanding of the structure and organization of transcriptional networks. (ii) Evolution: I then describe the different mechanisms and forces that influence network evolution and shape network structure. (iii) Dynamics: I discuss studies that have integrated information on dynamics such as mRNA abundance or half-life, with data on transcriptional network in order to elucidate general principles of regulatory network dynamics. In particular, I discuss how cell-to-cell variability in the expression level of TFs could permit differential utilization of the same underlying network by distinct members of a genetically identical cell population. Finally, I conclude by discussing open questions for future research and highlighting the implications for evolution, development, disease and applications such as genetic engineering.

Introduction
Regulation of gene expression at the transcriptional level is a fundamental mechanism that is evolutionarily conserved in all cellular systems [1]. This form of regulation is typically mediated by TFs (transcription factors) that bind to DNA and either activate or repress the expression of nearby genes [2,3]. Extensive research into understanding this process at the biochemical and structural level has provided us with knowledge that now allows us to investigate this process on a genomic scale (see Table 1 for genome-scale experimental strategies to probe protein–DNA interactions). These experimental approaches have been applied to several prokaryotes and eukaryotes, resulting in a wealth of information that is stored in publicly available databases [4–7]. To take advantage of this deluge of information, approaches that understand general principles on a genomic scale have become indispensable. In the present review, I provide an overview of the structure, evolution and dynamics of TRNs (transcriptional regulatory networks).

Structure and organization of the TRN
Experiments performed over the last several decades have resulted in a large amount of information on protein–DNA interactions and gene regulation in several model organisms [4–9]. In addition, advances in experimental techniques that detect protein–DNA interactions (see Table 1) have provided us with evidence for TF–DNA interactions on a genomic scale [10–19]. This information is best represented as the TRN with nodes connected by directed links (Figure 1a) [20,21]. In such a network representation, the nodes represent either TFs or TGs (target genes) and the directed links represent regulatory interactions (protein–DNA interactions) between TFs and TGs [21].

Key words: genome sequence, network dynamics, network evolution, network structure, transcriptional regulatory network.

Abbreviations used: ABF1, autonomously-replicating-sequence-binding factor 1; DOR, dense overlapping regulon; FFL, feedforward loop; FFM, feedforward motif; HGT, horizontal gene transfer; HNF, hepatocyte nuclear factor; Hns, histone-like nucleoid structuring protein; MIM, multiple-input motif; SIM, single-input motif; TE, transposable element; TF, transcription factor; TG, target gene; TRN, transcriptional regulatory network; WGD, whole genome duplication.

1 email madanmb@mrc-lmb.cam.ac.uk
Table 1 | Genome-scale experimental methods to probe protein–DNA interactions
Adapted from [85].

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ChIP-chip and ChIP-seq</td>
<td>The DNA-binding protein is tagged with an epitope and is expressed in a cell. The bound protein is covalently linked to DNA using an in vivo cross-linking agent such as formaldehyde. After cross-linking, DNA is sheared and the protein–DNA complex is pulled down using a tag-specific antibody. Reversal of the cross-link releases the bound DNA, allowing the sequence of the fragments to be determined by hybridization to a microarray (ChIP-chip) or by sequencing (ChIP-seq). In ChIP-chip experiments, intergenic regions are spotted on to a microarray chip. Following the ChIP step, the cross-links are reversed and the bound DNA is hybridized on to the chip. Complementary sequences will bind to specific spots on the chip, thereby providing an indication of the intergenic region to which the protein was bound [10,11,242]. In ChIP-seq experiments, the bound fragments are sequenced directly using 454/Solexa/Illumina sequencing technology. The sequences are then computationally mapped back to the genome sequence. Fragments that were bound by the protein will be sequenced several times providing a direct measure of enrichment of binding [243–245].</td>
</tr>
<tr>
<td>DamID [(Dam (DNA adenine methyltransferase) identification)]</td>
<td>To overcome any potential non-specific cross-linking of protein to DNA as could happen with ChIP-chip experiments, the DamID technique was introduced. The protein of interest is fused to an E. coli protein, Dam. Dam methylates the N6 position of the adenine in the sequence GATC, which occurs at reasonably high frequency in any genome (∼1 site in 256 bases). Upon binding DNA, the Dam protein preferentially methylates adenine in the vicinity of binding. Subsequently, the genomic DNA is digested by the DpnI and DpnII restriction enzymes that cleave within the non-methylated GATC sequence, and remove fragments that are not methylated. The remaining methylated fragments are amplified by selective PCR and quantified using a microarray [246].</td>
</tr>
<tr>
<td>PBMs (protein-binding universal DNA microarrays)</td>
<td>In contrast with the methods described above, this is an in vitro method to probe protein–DNA interactions. A DNA-binding protein of interest is epitope-tagged, purified and bound directly to a double-stranded DNA microarray spotted with a large number of potential binding sites. Labelling with fluorophore-conjugated antibody for the tag allows detection of binding sites from the significantly bound spots [247].</td>
</tr>
</tbody>
</table>

Representing transcriptional interactions as a network provides us with a conceptual framework and the right level of abstraction to uncover general principles of regulation on a genomic scale [20–22]. A number of recent studies on transcriptional networks of prokaryotes and eukaryotes have shown that the structure of such networks can be investigated in at least three distinct levels of organization [21]. At the most basic level, the network consists of a single regulatory interaction between a TF and its TG (Figure 1a). At the local level of organization, a number of studies have uncovered the fact that the basic unit is organized into units of transcriptional regulation, called network motifs (Figure 1b) [23]. Finally, at the global level of organization, the set of all transcriptional regulatory interactions in a cell has been shown to adopt an architecture that is hierarchical and scale-free (Figure 1c; see the ‘Global network structure’ section) [19,21,24–29]. Several computer programs are available to construct, visualize and analyse large biological networks, and some of the commonly used software is listed in Table 2.

Local network structure
A network motif is defined as a small pattern of interconnections that recur at many different parts of the network at frequencies much higher than what is expected by chance when compared with random networks of similar size [10,23,30]. Analysis of the transcriptional networks of Escherichia coli and Saccharomyces cerevisiae has revealed the presence of three commonly occurring motifs, each of which has distinct kinetic properties in the control of gene
expression [23], as follows. (i) FFM (feedforward motif) (Figure 1b, top), where a top-level TF regulates both the intermediate-level TF and the TGs, and the intermediate-level TF regulates the TG as well. If both TFs are activators, such a connectivity pattern might ensure that the TG is expressed only when persistent signal is received by the top-level TF. Since the concentration of the intermediate TF should be built up for the regulation of the final TG, random fluctuations and noise in the activation of the top-level TF are filtered and do not propagate. (ii) SIM (single-input motif) (Figure 1b, middle), where a single TF regulates the expression of several TGs simultaneously. Depending on the promoter strength of the regulated genes, it may respond to different concentrations of the active TF. Therefore, if the concentration of the TF changes with time, such a motif could set a temporal pattern in the expression of the individual targets. (iii) MIM (multiple-input motif) (Figure 1b, bottom), where multiple TFs regulate the expression of multiple TGs. Since the TFs could potentially respond to different signals, such motifs could therefore integrate diverse signals and bring about differential expression of the relevant targets. Thus regulation of genes via network motifs provides distinct ways of regulating gene expression. Since maintaining the right levels of TGs can influence fitness of an individual [31,32], it is very likely that controlling gene expression via distinct motifs is advantageous under different conditions [33]. Local network properties such as motif identification can be carried out using programs such as Mfinder, FanMod and Cytoscape (Table 2).

Global network structure
At the global level of organization, analysis of transcriptional networks has revealed that they display a hierarchical and scale-free topology [19,21,24–29]. Hierarchical structures are characterized by an organization where one observes multiple layers of regulators (i.e. some TFs regulate other TFs, which in turn regulate the TGs). Scale-free networks are characterized by the presence of a few highly influential TFs that regulate several genes and a large number of TFs that regulate only a few genes. The highly influential TFs are referred to as global regulators, or regulatory hubs, and their presence contributes to the inherent robustness of such a topology. In this case, robustness is defined as the ability of complex systems to function even when the structure of the system is perturbed significantly [34,35]. A scale-free topology is robust because random inactivation of genes will probably affect the TFs that regulate a few genes, as these occur in very high numbers. This would still leave a central highly connected subnetwork that may be functional. However, the downside of such a network structure is that they are vulnerable to targeted attacks of hubs, i.e. targeted removal of the very highly connected nodes will result in the collapse of the system into small sets of isolated fragments that no longer interact with each other. Therefore the global regulators are believed to be crucial for the robustness and functioning of the regulatory network [36]. In addition to the hierarchical and scale-free structure, several studies have shown that there is extensive combinatorial regulation of the TGs in the transcription network [27,37–42]. This essentially means that a gene is typically regulated by more than one TF. Such an extensive combinatorial regulation provides both regulatory plasticity and robustness by ensuring that genes are expressed at optimal levels in different conditions. Global network properties such as connectivity, hierarchy, modularity and clustering can be investigated using packages such as Cytoscape, Pajek and Topnet (Table 2).

Evolution of TRNs
The availability of completely sequenced genomes and the development of high-throughput experiments have facilitated investigation of the evolutionary history of genes from...
Table 2: Description of computational programs to analyse networks
Adapted from [86].

<table>
<thead>
<tr>
<th>Tool</th>
<th>Visualization</th>
<th>Drawing/ editing</th>
<th>Analysis</th>
<th>Network topology</th>
<th>Network motifs</th>
<th>Randomization</th>
<th>Graph comparison</th>
<th>Subnetworks</th>
<th>Neighbours</th>
<th>Cliques</th>
<th>Graph clustering</th>
<th>Motif discovery</th>
<th>Motif clustering</th>
<th>Website (URL)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graphviz</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://www.graphviz.org/">http://www.graphviz.org/</a></td>
<td>–</td>
</tr>
<tr>
<td>Tulip</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://tulip.labri.fr">http://tulip.labri.fr</a></td>
<td>[249]</td>
</tr>
<tr>
<td>BioLayout</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://www.biolayout.org/">http://www.biolayout.org/</a></td>
<td>[250]</td>
</tr>
<tr>
<td>Osprey</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://biodata.mshri.on.ca/osprey">http://biodata.mshri.on.ca/osprey</a></td>
<td>[251]</td>
</tr>
<tr>
<td>Pajek</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://vlado.fmf.uni-lj.si/pub/networks/pajek/">http://vlado.fmf.uni-lj.si/pub/networks/pajek/</a></td>
<td>[252]</td>
</tr>
<tr>
<td>Yed</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://www.yworks.com">http://www.yworks.com</a></td>
<td>–</td>
</tr>
<tr>
<td>NetMiner</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://www.netminer.com">http://www.netminer.com</a></td>
<td>–</td>
</tr>
<tr>
<td>WANTED</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://vanted.ipk-gatersleben.de/">http://vanted.ipk-gatersleben.de/</a></td>
<td>[253]</td>
</tr>
<tr>
<td>VisANT</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td>√</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://visant.ubu.edu/">http://visant.ubu.edu/</a></td>
<td>[254]</td>
</tr>
<tr>
<td>Bioconductor</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://www.bioconductor.org">http://www.bioconductor.org</a></td>
<td>[255]</td>
</tr>
<tr>
<td>igraph</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://igraph.sourceforge.net/">http://igraph.sourceforge.net/</a></td>
<td>–</td>
</tr>
<tr>
<td>Cytoscape</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://www.cytoscape.org">http://www.cytoscape.org</a></td>
<td>[256]</td>
</tr>
<tr>
<td>NeAT</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://rsat.scmb.uib.ac.be/neat/">http://rsat.scmb.uib.ac.be/neat/</a></td>
<td>[257]</td>
</tr>
<tr>
<td>TYNA/Topnet</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://tyna.genestalab.org/tyna/">http://tyna.genestalab.org/tyna/</a></td>
<td>[258]</td>
</tr>
<tr>
<td>MCODE</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://baderlab.org/Software/MCODE">http://baderlab.org/Software/MCODE</a></td>
<td>[259]</td>
</tr>
<tr>
<td>NetMatch</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://baderlab.org/Software/NetMatch">http://baderlab.org/Software/NetMatch</a></td>
<td>[260]</td>
</tr>
<tr>
<td>MAVisto</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://msi.vist.nwu.edu/ipk-gatersleben.de/">http://msi.vist.nwu.edu/ipk-gatersleben.de/</a></td>
<td>[261]</td>
</tr>
<tr>
<td>Mfinder</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://www.weizmann.ac.il/mcb/">http://www.weizmann.ac.il/mcb/</a> UniAlon/groupnetworkMotifsW.html</td>
<td>–</td>
</tr>
<tr>
<td>FanMod</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://www.minet.uni-jena.de/~wernicke/motifs/">http://www.minet.uni-jena.de/~wernicke/motifs/</a></td>
<td>[262]</td>
</tr>
<tr>
<td>Mfinder</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://www.weizmann.ac.il/mcb/">http://www.weizmann.ac.il/mcb/</a> UniAlon/groupnetworkMotifsW.html</td>
<td>–</td>
</tr>
<tr>
<td>CFinder</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://www.cfinder.org/">http://www.cfinder.org/</a></td>
<td>[263]</td>
</tr>
<tr>
<td>Clique finder</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://topnet.genestalab.org/clique/">http://topnet.genestalab.org/clique/</a></td>
<td>[264]</td>
</tr>
<tr>
<td>BIOPROSPECTOR</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://bioprospector.stanford.edu/">http://bioprospector.stanford.edu/</a></td>
<td>[265]</td>
</tr>
<tr>
<td>CISMODULE</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://www.stat.stanford.edu/~zhou/">http://www.stat.stanford.edu/~zhou/</a></td>
<td>[266]</td>
</tr>
<tr>
<td>EMCMODULE</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CisModule/</td>
<td>[267]</td>
</tr>
</tbody>
</table>
different organisms. This has allowed us to gain insights into the interplay of evolutionary forces that drive the evolution of TRNs. In this section, I first provide a short overview of the major mechanisms of gene evolution and then discuss the role of these evolutionary forces in shaping TRNs.

**Mechanisms for the evolution of TRNs**

Mutations in the genome of an organism contribute to the evolution of TRNs. Such mutations may affect just a single or few bases (e.g. single nucleotide substitutions) or may result in the generation of a large chunk of genetic material (e.g. duplication, repeat element expansion by transposons or horizontal transfer). Accordingly, such events may have a range of outcomes; for instance, they can affect regulatory interactions either (i) at the level of cis-regulatory elements, by introducing point mutations in cis-regulatory elements or incorporating new cis-regulatory elements upstream of genes during repeat element expansion, or (ii) at the level of trans-factors, through the modification of pre-existing or generation of new DNA-binding domains that may recognize a different DNA sequence or may respond to a different ligand. Most of these mutations are likely to be either deleterious or cause disruption of an existing regulatory interaction. Evolution of the TRNs, on the other hand, requires the addition of new nodes (TFs and TGs) and new edges (regulatory interactions). As discussed in the following sections, gain of genes is crucial for these two aspects. As illustrated in Figure 2, gene gain is driven either by gene duplication [21,43,44] or by HGT (horizontal gene transfer) (primarily in prokaryotes) [45,46]. Whereas these two processes intrinsically add new nodes in TRNs, more importantly, they increase the evolvability of such networks by facilitating gain and rewiring of regulatory interactions during divergence [21,47–49]. This point is well illustrated by a recent study which showed that artificial incorporation of new regulatory interactions into *E. coli* is rarely a barrier for evolution and even contributes to the fitness under various selection pressures [50]. In the first part of this section, I primarily consider gene duplication, gene loss and HGT and then briefly describe evolution of new interactions through repeat element expansion in the latter part.

**Impact of gene duplication on TRN evolution**

Evolution by gene duplication involves the generation of a second copy of the genomic segment containing a gene, thereby resulting in two identical copies of the same gene in an organism. Following duplication, one of the copies retains the ancestral function and the other copy may diverge under a relaxed selection pressure until it acquires a new function (neo-functionalization). Alternatively, the two copies may share a part of the function of the ancestral copy (sub-functionalization) or the second copy may degenerate into a pseudogene [51]. In a simplistic scenario, three different cases (Figure 2) must be considered: i.e. whether the duplicated segment contains either a TF or a TG, or both [21,52,53]. As an immediate consequence of this event, gene duplication will result in doubling the quantity of regulatory interactions
in addition to the number of genes involved. In each case, the fate of those shared interactions, i.e. their maintenance or removal during evolution, is of importance to understand the evolution of transcriptional networks.

Through a systematic analysis of the TRN of the prokaryote *E. coli* and the unicellular eukaryote *S. cerevisiae*, we found that more than two-thirds of the interactions have evolved as a consequence of gene duplication [53]. We also observed that over one-half of the known regulatory interactions were inherited from ancestral TFs or TGs after duplication. The remaining interactions seem to have evolved as a consequence of rewiring (i.e. loss and gain of interactions) or gain of new interactions during divergence after gene duplication [21,52,53]. We also noticed that only a small fraction of the genes and the regulatory interactions have evolved as a consequence of gene recombination or innovation [21,53].

An obvious question that arises given the vast amount of gene duplication during the evolution of transcription networks is whether this has had any significant role in the generation of the network motifs or of the global topology of the network. In the same study [53], we demonstrated that, although individual genes in network motifs may have evolved as a consequence of gene duplication, the interactions that make up the motifs have either been gained or have evolved as a consequence of rewiring. Conant and Wagner [54] also observed the same trend by investigating the yeast and the *E. coli* network. These studies together demonstrate that network motifs have evolved independently (i.e. convergent evolution) multiple times, possibly because they contribute to fitness by tuning the expression level of genes in a way that maximizes fitness. This is supported by the observation from experimental evolution studies, where *E. coli* was found to optimize its expression level of a protein that maximizes growth rate and therefore its fitness [31]. An investigation of the global structure of the TRN showed that the scale-free topology is not a direct consequence of gene duplication [53]. Whereas this observation is consistent with the possibility that the scale-free structure could have evolved due to selection, there are other possible mechanisms, which are non-adaptive (e.g. neutral evolution), that may also give rise to the same scale-free-like topology [55].

Although duplication of individual genes is one mechanism to create new genes, WGD (whole-genome duplication) is another mechanism that results in the generation of copies of multiple genes simultaneously. It is now well accepted that this mechanism has contributed to the evolution of genomes and their regulatory networks. Several studies [56–59] have investigated the role of WGD on network evolution. Such studies have revealed that (i) genes targeted by many TFs appear to be preferentially retained in duplicate after WGD, (ii) duplicate TFs produced by WGD rapidly evolve to perform distinct functional roles in the regulatory network, (iii) newly formed transcriptional pathways remain connected (paths are not broken) and are preferentially cross-connected with ancestral ones, and (iv) duplicated TFs through WGD may have contributed to the evolution of some network motifs in yeast. Taken together, these studies have
shown that gene duplication has played a key role in the evolution of the network components, particularly losses, gain and rewiring of regulatory interactions, thereby fuelling network evolution.

**Impact of HGT on TRN evolution**

In eukaryotes, gene duplication and loss are believed to be the major source of genome diversification. However, in prokaryotes, horizontal transfer of genetic material represents a substantial source of genetic novelty [45,60]. HGT requires the physical incorporation of foreign DNA into the receiver organism, its integration into the host regulatory network and eventually its selection through the bacterial population (i.e. its fixation). The incorporation of DNA during HGT is driven by three distinct mechanisms referred to as conjugation, transduction and transformation [61]. Interestingly, the uptake of foreign genes is often biased towards the acquisition of traits that contribute directly to fitness such as virulence, symbiosis or resistance to toxins [62–64]. Thus, while understanding that the role of HGT is of particular importance in prokaryotic evolution, it also has implications for understanding how foreign genes contribute to network evolution and adaptation of organisms to new environments [65,66]. In this section, I discuss the regulatory constraints and mechanisms that shape the integration of new genes in TRNs. When a segment of DNA is horizontally transferred into an individual, the immediate impact of the imported genes on fitness is indeed crucial for the adaptation and survival of the individual in a bacterial population. However, how the gene becomes integrated into the chromosome and how it integrates into an existing regulatory network is only now being understood in detail [67–71] (Figure 2).

If the transferred segment is transcriptionally active, the imported gene must be successfully translated and should not be lethal when it is expressed. In such cases, its protein expression level must be adequately regulated. This implies the need for a tighter transcriptional regulation, and thus a proper recognition of its promoter region and TF-binding sites by the host transcriptional network. Otherwise, it might require a horizontally transferred TF that came along with the DNA segment. Therefore the probability of integrating a transferred gene into a network is expected to generally decrease with phylogenetic distance [64]. It has been observed in *E. coli* K-12 that genes in K-loops, known to be hotspots of HGT, are poorly translated [72]. Taoka et al. [72] provided evidence that most of the recently acquired foreign genes in *E. coli* K-12 are generally not translated under laboratory conditions, suggesting that their expression may not directly contribute to fitness (i.e. growth) in exponential-phase culture. In another study, Sorek et al. [64] have shown that genes that failed to be horizontally transferred are those that are generally highly expressed. Thus viability and successful synthesis of newly acquired genes alone are unlikely to be sufficient conditions for fixation. A balance between fitness benefits and cost of synthesis of the new gene is therefore necessary for the survival and competitiveness of the individuals harbouring the transferred gene in a genetically mixed bacterial population.

How can the cell find a strategy to favour such balance? Interestingly, several recent reports have suggested that it might be important, as a first step, to silence the transferred gene. The transferred gene can then be subsequently expressed (through anti-silencing mechanisms [69]) when the benefit of its expression is higher than the cost of its synthesis. This is likely to tip the balance in the population, favouring the emergence of individuals who harbour the transferred gene. For example, it was observed that nucleoid-associated proteins such as Hns (histone-like nucleoid structuring protein) contribute to silencing the transcriptional activation of recently acquired genes. This mechanism provides a ‘stealth function’, minimizes the cost on fitness of their expression and thus facilitates their transmission [69,73]. Consistently, Navarre et al. [70,74] and Lucchini et al. [75] demonstrated that, in *Salmonella*, Hns selectively silences horizontally acquired genes by targeting sequences with a GC-content lower than that of the resident genome. In addition to these studies, Perez and Groisman [76] have suggested that mutations in orthologous TFs and in their dependent promoters in different organisms may allow bacterial TFs to incorporate newly acquired genes into ancestral regulatory circuits and yet retain control of the core members of a regulon (i.e. the set of genes regulated by a TF). In this context, Cosenzino Lagomarsino et al. [29] have shown that horizontally transferred genes are mainly added at the bottom of the hierarchy of the existing gene-regulatory network, primarily as new target nodes. Taken together, these studies have begun to help us understand the role of HGT in network evolution and to better appreciate various aspects of laterally acquired genes that contribute to their increased likelihood of being successfully integrated into existing regulatory networks.

**Impact of TEs (transposable elements) on TRN evolution**

In addition to the above discussed mechanisms, there is increasing evidence on a genomic scale that TEs contribute to regulatory network evolution in eukaryotes in two ways: (i) they contribute DNA-binding domains that are recruited as TFs in the host genome, and (ii) they provide additional cis-regulatory elements by carrying them during repeat element expansion, thereby introducing new TF–DNA interactions in the host genome. In line with the possibility that some TFs might have evolved from selfish elements, we presented evidence that the WRKY family of DNA-binding domains, which are seen in TFs from several eukaryotic lineages, evolved from the MULE (Mutator-like element) transposases [77]. Upon recruitment as a TF by the host, they independently underwent lineage-specific expansion in the different eukaryotic lineages, thereby contributing to the evolution of the TRN in these organisms. Similarly, in another study where we reported the discovery of a novel family of lineage-specifically expanded TFs containing the AP2 (Apetala2) DNA-binding domain, we presented evidence that
the DNA-binding domain was associated with mobile selfish replicative elements [78].

A number of studies have provided convincing evidence that selfish elements donate *cis*-regulatory elements, thereby introducing new regulatory interactions in an organism [79]. Jordan et al. [80] investigated human gene promoters and revealed that almost 25% of the promoter regions analysed contain TE-derived sequences, including many experimentally characterized *cis*-regulatory elements. They also showed that S/MARs (scaffold/matrix-attachment regions) and LCRs (locus control regions) that are involved in the simultaneous regulation of multiple genes also contain numerous TE-derived sequences [80]. Bourque et al. [81] investigated the binding profile of a number of evolutionarily conserved TFs such as Sox2 or TP53 and showed that the binding sites for several of these factors were embedded within repeat elements. Wang et al. [82] showed that several targets of the human proto-oncogene TF c-Myc have evolved from expansion of the TE families L2 and MIR. In addition, they also showed that the subset of genes regulated by TE-derived c-Myc-binding sites appear to form a distinct c-Myc regulatory subnetwork in terms of their expression profile. Consistent with these studies, Xie et al. [83] and Kunarso et al. [84] showed that the transcriptional network that governs embryonic development and the network of interactions involving Oct4 and Nanog respectively evolve rapidly between mammalian genomes. More importantly, they showed that this rapid rewiring is primarily due to repeat element expansion. Taken together, these studies clearly demonstrate that TRNs are dynamic in eukaryotic genomes and that TEs play an important role in expanding the repertoire of TFs and binding sites in a genome.

**Evolution of networks across organisms**

Although the above studies have provided insights into how networks evolve within an organism, it is of fundamental interest to understand how TRNs evolve across species. In other words, are interactions between TFs and TGs sufficiently conserved to be able to predict a regulatory interaction in an organism from a closely related one? This question is important since less information is available on the transcriptional networks of many organisms, as most of the experimental studies performed over the last few decades have focused on model organisms such as *E. coli*, *Bacillus subtilis* and *S. cerevisiae*. Approaches used to address the problem of the inference of TRNs from other prokaryotes can broadly be grouped into two categories, depending on whether one focuses on orthology or on sequence similarity of TF-binding sites [85–88]. The first category of methods exploits the assumption that orthologous TFs regulate orthologous TGs in distinct genomes. The latter set of methods exploits the assumption that identical binding sites upstream of two genes in closely related species imply similar regulatory interactions with orthologous TFs. Altogether, these methods have provided us with a deeper insight into the evolution of TRNs across organisms.

Studies by us [33], Lozada-Chavez et al. [89] and Price et al. [90] investigated over 150 completely sequenced genomes and have shown that TFs are less conserved across genomes than their TGs, suggesting a greater evolvability of TFs. Noticeably, it was observed that global regulators do not differ from other specific TFs in terms of sequence conservation. Another study by Hershberg and Margalit [91] showed that the mode of regulation (activation or repression) exerted by TFs has an effect on their evolution. Repressors were found to co-evolve tightly with their TGs. In contrast, activators were found to be lost independently of their targets. These results suggest that prokaryote organisms rapidly evolve their own set of transcriptional regulators, and are therefore able to rewire regulatory interactions in a very flexible way. These observations are also supported by a study by Isalan et al. [50], which has shown that artificial incorporation of new regulatory interactions into *E. coli* is not a barrier for evolution and in fact contributes to increased fitness under certain selection pressures.

An analysis of the local structure revealed that motifs are not conserved as whole units and that individual interactions within a motif may be lost or retained. Given the functional importance of network motifs, these results may seem surprising at a first glance, as one would have expected that closely related species will conserve local network structures. However, we have shown that organisms that live in similar environments tend to conserve similar interactions and motifs [33]. In fact, it was noticed that losing or gaining interactions can result in embedding orthologous genes in different motif contexts (Figure 2). This result is more meaningful when one considers the environment in which an organism lives. This trend appeared to be statistically significant and we identified a number of interesting examples [33]. For instance, in *E. coli*, it was observed that the fumarate reductase genes FrdB and FrdC are under the control of the TFs Fnr and NarL in an FFM. These enzymes, which convert fumarate into succinate under anaerobic conditions to derive energy, are therefore only expressed when both Fnr and NarL are active, which only occurs under a persistent signal due to lack of oxygen. Consistently, *E. coli* faces alternations of aerobic and anaerobic phases over long periods, which makes it important to induce fumarate reductases only when the bacterium is likely to stay in an anaerobic environment for extended periods of time. In contrast, *Haemophilus influenzae* is a pathogen that faces strong redox fluctuations during host infection when it passes through the arteries and veins. Interestingly, in contrast with what happens in *E. coli*, NarL is lost, and the expression of FrdB or FrdC only depends on Fnr. Therefore the fumarate reductases are regulated in a simpler manner (through a SIM) in this pathogen, which again seems relevant given its environmental lifestyle. Interestingly, this FFM found in *E. coli* is also conserved in distantly related organisms, such as *Bordetella pertussis* (*β*-proteobacterium) and *Desulfotobacterium hafniense* (firmicute), which live in similar environments. At the level of the global structure,
it was observed that global regulatory hubs are not preferentially more conserved than other TFs (i.e. non-hubs). It was found that the condition-specific global regulatory hubs are the ones that may be lost more easily. This observation lends support to the idea that orthologous TFs may contribute to different extents to the fitness of organisms living in different environments, and hence completely different TFs may emerge as global regulators. Consistent with this, an analysis of the E. coli and the B. subtilis networks revealed that, although the global topology was similar, very different proteins emerged as global hubs. This observation again points to the importance of the environment in shaping network structure [33].

Taken together, these observations highlight an important principle: TRNs are extremely plastic, evolve rapidly and adapt to the environment by tinkering with individual interactions [33,89,90,92]. More specifically, the general principles can be summarized as follows (Figure 3): at the level of network components, TFs evolve more rapidly than their TGs, allowing organisms to evolve their own set of regulators in line with their environment. Besides, at both the basic and the local structure levels, organisms with similar lifestyles tend to possess similar regulatory interactions and motifs. Finally, at the level of the global structure, conservation of TFs is independent of their connectivity (i.e. the number of TGs), whereas the environment, again, seems to be the major force driving gain and loss of TFs and regulatory interactions.

In addition to studies in prokaryotes, several experimental studies in a number of eukaryotes have shown that the TGs of orthologous TFs tend to diverge rapidly among related species [83,84,93–99]. Odom et al. [96] mapped the binding sites of four conserved TFs [FOXA2 (forkhead box A2), HNF (hepatocyte nuclear factor) 1A, HNF4A and HNF6] involved in liver development in humans and mouse. They showed that between 41 and 89% of the binding events are species-specific, suggesting rapid evolution of regulatory interactions in conserved TFs. A recent, more extensive, study by Schmidt et al. [97] compared the binding of two conserved TFs [C/EBPα (CCAAT/enhancer-binding protein α) and HNF4α] among five vertebrate genomes and showed that most binding events were species-specific. Borneman et al. [94] performed ChIP-chip (chromatin immunoprecipitation on chip) experiments on the TFs Ste12 and Tec1 in S. cerevisiae, Saccharomyces mikatae and Saccharomyces bayanus under pseudohyphal conditions and showed that several sites (to which these TFs bound) were different between the three species. Similarly, Lavoie et al. [93] investigated the binding profiles of a number of conserved TFs that regulate ribosomal genes in S. cerevisiae and Candida albicans and demonstrated significant differences in their binding profiles. More importantly, they demonstrated that orthologous TFs have been recruited to regulate genes involved in different functions (Figure 4).

A recent study by Kunarso et al. [84] analysed the binding profiles of two factors critical for stem cell maintenance, namely Oct4 and Nanog, from humans and mouse and showed that (i) only 5% of the regions are homologously occupied, and (ii) TFs, which carried cis-regulatory sequences for these factors, contributed to the uniquely bound sites in humans and mouse. Likewise, Xie et al. [83] compared the bovine, murine and human transcriptional network that governs embryonic development and showed that there have been extensive changes in regulatory interactions mediated by single nucleotide changes and TEs affecting cis-regulatory elements. They also present evidence for changes in regulatory interactions due to alterations in trans, resulting in a new combinatorial interaction and associated regulatory interactions with such a change. The latter two studies also suggest that transcriptional circuitry of pluripotent stem cells have been rewired by TEs. Although all the above studies demonstrate differences in the binding profiles, they also report a number of binding events that are evolutionarily conserved. This suggests the existence of core regulatory networks that are evolutionarily conserved, with several new regulatory interactions that are species-specific and may contribute to the characteristic phenotype of an organism [100].

Recent exciting studies have also shown that transcriptional interactions diverge even between individuals in a population or among different strains of the same species [14,15]. In this context, Kasowski et al. [15] compared the binding of RNA polymerase II and the TF p65 in ten lymphoblastoid cell lines obtained from different human individuals and showed that 25 and 7.5% of the respective binding sites are different between individuals. They implicated single nucleotide changes and structural alterations, such as copy number changes, insertions and deletions, as causative factors for this divergence in the binding events between the different individuals. Similarly, Zheng et al. [14] investigated the binding profile of Ste12 from 43 different segregants obtained using two divergent yeast strains as parents. They showed that binding events are strikingly different between the different individuals and that such differences are likely to be functional in terms of introducing changes in the levels of gene expression of the TGs. Taken together, these studies suggest that, during the evolution of eukaryotes, there have been extensive changes, which occurred in TF wiring, DNA-binding specificity and combinatorial regulation (Figure 4). It also shows that TRNs evolve rapidly in prokaryotes, unicellular eukaryotes and multicellular organisms.

**Dynamics of TRNs**

Cells continuously sense and respond to environmental changes and internal signals in a highly noisy environment. This is possible due to a complex cascade of interactions of biomolecules with the TFs, which ultimately controls the state of the cell by influencing the expression of relevant genes. The interplay between cis-regulatory elements and TFs provides a plethora of transcriptional programs, which regulate the state of every gene in the cell tailored for different conditions. Despite several studies which focus on regulatory networks as a static entity, it should be noted that the
regulatory network of an organism is highly dynamic and different sections of the network are used under different conditions [101–104]. Although informative, the topological properties of regulatory interactions themselves explain little about how a TRN functions. To understand the principles governing the logic behind how various parts of the static network operate dynamically in a given condition, one needs to investigate changes in the network structure of an organism across different conditions at a global level and at the level of individual regulatory interactions. Accordingly, recent studies that attempted to understand the dynamics of gene regulation can be split into three distinct categories: (i) those that concentrate on understanding the large scale changes in gene-regulatory interactions across different cellular conditions or over a period of time, (ii) those that focus on identifying the principles governing the functioning of local structures such as motifs or circuits in the context of a larger system, and (iii) those that investigate the dynamics of transcription and translation of individual genes that make up the network.

**Temporal dynamics of transcriptional networks**

Cells have the ability to respond to most dynamic changes, whether simple (involving the changes in exogenous conditions such as variations in temperature or nutrient concentrations) or complex (involving the simultaneous change of many conditions). In order to accurately process and respond to complex environmental changes, organisms use distinct transcriptional regulatory subnetworks, governed by sets of sensors which are specialized to detect environmental stimuli [105,106]. By integrating gene-expression data across five different conditions with the static TRN of yeast, we showed that the active subnetwork for the different conditions vary significantly in terms of both their local and global structure (Figure 5a) [101]. The five conditions were cell cycle, sporulation (both of which are developmental regulatory programmes in a cell), DNA damage, stress response and diauxic shift (all three of which are regulatory programmes that are important for survival). We identified the condition-specific transcriptional network by linking TFs present in a given condition to their differentially expressed TGs. TFs were classified as ‘present’ or ‘absent’ on the basis of their abundance in a given condition. Although a TF might bind physically to its target site, the corresponding link was not considered active if the expression of the TG did not change significantly, or if the TF abundance was low under the specific condition. Analyses of the resulting subnetworks revealed that the majority of regulatory interactions are condition-specific, and only a small subset is active in four or more conditions (Figure 5b). Importantly, we found that most hubs are transient and their expression is not maintained between conditions. Only a small percentage of hubs were found to regulate many genes in all conditions. However, even these hubs, which maintained a large number of regulated TGs across conditions, were found to switch their targets between different conditions. As a result of extensive rewiring, the same TFs can be used in different conditions to regulate the expression of various sets of genes and to elicit a condition-dependent response. This implies that the response of a
cell is commonly a result of combinatorial TF usage. Taken together, we suggested that the transcriptional networks are extremely dynamic and rapidly rewire their interactions to respond efficiently to changes in the external and internal environment.

In another genome-wide study in yeast, Harbison et al. [104] determined the location of TF-binding sites in the promoter regions for 203 TFs in rich medium. In addition, a genome-wide binding profile was obtained for 84 TFs in at least one of the 12 different environmental conditions. By obtaining the transcriptional network, they addressed questions on the organization of regulatory elements in promoters in the context of their environment-dependent usage. Through the compilation of information from TF-binding data, evolutionarily conserved sequences and prior knowledge, they inferred a detailed TRN for 102 TFs. In contrast with the previous approach, this approach experimentally identified all promoters that are bound by TFs in the different conditions. On the basis of the presence of TF-binding sites, the authors classified the identified promoter regions into four different architectures, called ‘single regulator’, ‘repetitive motif’, ‘multiple regulators’ and ‘co-occurring regulator’ types (Figure 5c). The first two of these architectures are characterized by the presence of one or more binding sites for a single TF, whereas the other two contain binding sites for two or more TFs. They also classified promoter utilization patterns into four types: ‘condition invariant’, in which the set of TF-binding sites does not change across conditions; ‘condition enabled’, in which the TF binds in one condition but not in the other; ‘condition expanded’, in which the set of binding sites in one condition includes those used in the other; and ‘condition altered’, in which different sets of promoters are bound in the two conditions (Figure 5d). In summary, the above two studies highlight that the model eukaryote S. cerevisiae uses largely distinct parts of its regulatory network in different environmental conditions via combinatorial usage of TFs in the promoter region of the relevant TGs.
A number of studies have also demonstrated that prokaryotic regulatory networks are as dynamic as the eukaryotic network [103,107–109]. Balazsi et al. [107] integrated gene expression data with the regulatory network in *E. coli* and systematically identified topological units called ‘origons’, which are regulatory subnetworks that originate from a TF that can sense a signal. Such an approach basically assumes that different subnetworks from the completely known TRN could be active under different environmental conditions depending on the signals sensed by the sensor TFs. In another study, Balazsi et al. [108] investigated the dynamics of gene networks in *Mycobacterium tuberculosis* associated with the switch between replicative (growth) and non-replicative (dormancy) states in response to host immunity. They integrated microarray data after assembling the transcriptional network from published literature. It was shown that distinct sets of transcriptional subnetworks (origons) were responsive at various stages of adaptation, revealing a gradual progression of network response under both conditions. Most of the responsive origons were found to be common between the two conditions, thereby defining a general transcriptional signature of *M. tuberculosis* growth arrest.

In another study, Martinez-Antonio et al. [103] integrated functional data with the transcriptional network of *E. coli* and showed that short pathways, multiple feedforward loops and negative auto-regulatory interactions are particularly predominant in the subnetworks that control metabolic functions such as the use of alternative carbon sources. In contrast, long hierarchical cascades and positive auto-regulatory loops are overrepresented in the subnetworks that control developmental processes such as biofilm formation and chemotaxis. Marr et al. [109] integrated information about gene expression with the transcriptional network in
E. coli and defined two different modes of gene regulation, namely the digital and analogue type. The digital mode represents dedicated TFs that bind at high affinity and regulate a set of genes in a specific manner, whereas the analogue mode consists of DNA architectural proteins which bind at low affinity, modulate superhelicity and affect expression of several genes in a less specific manner. Taken together, these studies are beginning to reveal general principles of the large-scale dynamics of TRNs across different conditions in prokaryotes and eukaryotes.

**Dynamics of individual regulatory interactions and network motifs**

As discussed above, the basic unit in a transcriptional network comprises a regulatory interaction between a TF and its regulated gene. However, an even simpler system would be if the TF regulated its own expression, i.e. the TG would be the TF itself. This mode of regulation is referred to as auto-regulation. In real regulatory networks, TFs are known to be both negatively and positively auto-regulated with preference for negative auto-regulation in several experimentally known TFs. Negative auto-regulation occurs when a TF represses its own transcription. This simple circuit is known to have two important functions: (i) speeding up the response time, and (ii) to reduce the cell-to-cell variation in protein levels [110–112]. Similarly, positive auto-regulation occurs when a TF activates its own transcription by up-regulating it. A positive auto-regulatory circuit has been shown to have effects which are opposite to what is seen for negative auto-regulatory feedback loops [113]. Indeed, strong positive auto-regulation has been shown to lead to extensive variations in protein levels among cells and is believed to lead to a differentiation-like partitioning of cells in a population [114]. Therefore this circuit is proposed to help cell populations maintain a mixed phenotype in order to survive when exposed to a mixture of external conditions.

In addition to auto-regulatory motifs, network motifs are known to occur frequently in transcriptional networks and have been shown to possess distinct dynamic and kinetic properties (see the ‘Local network structure’ section). The FFL (feedforward loop) is one of the most characterized multi-gene motifs in terms of dynamics. As discussed in the ‘Local network structure’ section, out of the eight different subtypes of FFLs, coherent type 1 and incoherent type 1 have been shown to be the most abundant in real transcriptional networks [23]. Each of the coherent and incoherent types of FFLs can have an AND or OR input function at the promoter of the TG, depending upon whether both or only one of the two TFs are needed to regulate the target. Coherent FFL with AND input has been shown to be responsible for an initial delay in the transcription of the TGs involved in the arabinose utilization system in E. coli [115]. Since the input function is AND, any sporadic signals are filtered out by this motif and only persistent signals result in the expression of the TG. On the other hand, the same motif with OR input shows an initial quick response, but exhibits a lag when the signal is shut down, as has been demonstrated in the flagellar system of E. coli [116]. In this study, the FFL regulating expression of the flagellar genes was found to prolong their expression even after the input signal stopped. However, no delay was detected when the input signal appeared. In an incoherent FFL, the effect of regulation from the two TFs involved is opposite. This motif gives rise to a pulse-like behaviour as the onset of the expression of the TF directly controlling the TG can lead to its sudden increase/decrease of expression. However, once the second TF reaches its repression/activation threshold with some time lag, it can lead to a complementary effect on the transcription of the target. This motif structure has also been shown to act as a response accelerator due to the capacity to bring about sudden changes in the expression of the TG [117].

The dynamics of SIMs, in which a single TF regulates several genes, have been well characterized in recent years. This motif has been shown to introduce a temporal order in the regulated genes, thereby providing just-in-time transcription of the set of TGs of a TF. Typically in this motif, the TF regulating different genes has a different activation/repression threshold for each gene. Therefore, when the concentration of the TF increases, it crosses the thresholds in a defined order, following a temporal order of expression of the TGs. This motif has been studied in great depth in the case of metabolic pathways such as the arginine biosynthesis pathway in E. coli where the arginine repressor is known to regulate several operons involved in the biosynthesis of arginine. It was found that when arginine is removed from the system, promoters of the enzymes responsible for its biosynthesis are activated in a temporal order just in the order needed for its biosynthesis [118].

Although other motifs such as MIMs or DORs (dense overlapping regulons) are known to be enriched in transcriptional networks, the difficulty in collecting experimental data due to their complex topology has limited our understanding. A study by Kaplan et al. [119] investigated parts of the DOR, which comprises five different sugar utilization systems in E. coli, and found that diverse and intricately shaped input functions governed the different sugar utilization systems, despite the similarity in the regulatory circuitry of these genes.

**Dynamics of networks in individual members in a cell population**

Although transcriptional regulatory interactions are conveniently represented as nodes and edges in a network, it should be noted that each node in the network represents several entities (gene, mRNA and protein) and events (transcription, translation, post-translational modification, degradation, etc.) that are compressed in both space and time. Consequently, the dynamic nature of these events (synthesis and degradation of mRNA and protein molecules) and entities (steady-state levels of mRNA and protein molecules) are expected to affect the regulatory interactions in the network. Although we have a good understanding of the topology of regulatory networks, the dynamics of nodes (TFs and TGs) in these networks and their role in systems behaviour remain largely unexplored. In this regard, several
fundamental questions remain to be answered: for example, do TFs in the regulatory network have distinct dynamic properties (e.g. abundance or half-life) that characterize their role in a regulatory cascade? More generally, does the position of a TF in the network structure relate to its dynamics? Although the richness of this detail is lost in the network representation, such questions can be addressed by integrating diverse genomic datasets encapsulating the dynamics of transcription and translation.

In a recent study [27], we investigated the dynamics of the yeast TFs by integrating diverse genome-scale datasets with the inherent hierarchical structure in the yeast transcription regulatory network. We reported a new graph-theoretical algorithm called vertex sort, to elucidate the precise topological ordering of vertices (nodes) in any directed network, acyclic or cyclic. Applying vertex sort on the yeast regulatory network revealed the presence of seven hierarchical levels, which could be grouped into three mutually exclusive hierarchical layers of TFs: top, core and bottom. We observed that the network is organized in such a way that its architecture is in between a strictly hierarchical ‘autocratic’ structure (with multiple hierarchical levels and no single giant strongly connected component) and a highly interconnected ‘democratic’ structure (in which a few master-regulator TFs regulate the set of other TFs that mutually regulate each other, effectively forming a two-level hierarchy). This observation of a multi-level hierarchy with a strongly connected component (core-layer; strongly connected component) is consistent with what has been discussed recently by Bar-Yam et al. [120]. In this direction, a recent study by Bhardwaj et al. [121] reports methods to quantify hierarchical structures and the extent of democratic or autocratic structure in a more quantitative manner.

We then integrated multiple genome-scale datasets with the hierarchical structure of the yeast regulatory network and observed that transcript half-lives of TFs from the three layers are comparable, but that the top-layer TFs are present in relatively higher abundance at the protein level and have a much longer protein half-life when compared with that of core- and bottom-layer TFs. This suggests that post-translational regulation plays an important role in ensuring the availability of appropriate amounts of each TF within the cell. The need for the presence of top-layer TFs to relay faithful signals down the transcriptional cascade could explain why top-layer TFs need to occur in relatively higher abundance than the core- and bottom-layer TFs. These findings are supported with what has been proposed by Farkas et al. [122] who suggest a model in which regulatory cascades originating from distinct fractions of the regulatory network control robust integrated responses to complex stimuli.

We also observed that top-layer TFs display a relatively higher variability in protein abundance between individuals in a clonal population of cells. This is an important observation as it suggests that such behaviour may confer a selective advantage on individuals by permitting at least some members in a population to respond effectively to changing conditions by triggering relevant transcriptional cascades [123–136]. For example, we observed that ABF1 (autonomously-replicating-sequence-binding factor 1), a multifunctional TF present in the top layer, is an abundant protein whose levels are noisy in a clonal population of cells. However, the activity of ABF1 depends on the availability of its co-activators (e.g. Cdc6) and on its phosphorylation state, which is known to be regulated by several kinases (e.g. casein kinase 2) or phosphatases [137,138]. The relatively higher noise in the abundance of ABF1 was proposed to ensure that at least some members in a population would respond rapidly during changing environments (i.e. when co-activators or kinases are activated in response to the altered external stimulus). We proposed that high variability in the expression of key TFs, whose TGs might contribute to phenotypic variation, might be a general strategy to facilitate adaptation to diverse environments (see Figure 6 for a model). This does not exclude the possibility that variation in the protein expression levels of specific TGs (independent of the variation in the levels of their regulating TFs) might dictate cell-fate outcomes in a post-transcriptional or post-translational manner. Hence, a detailed investigation that integrates multiple types of networks with data on cell-to-cell variation in transcript and protein levels might elucidate the contribution from TF-dependent and TF-independent modes for adaptability of cells to changing environments.

Furthermore, we observed that the protein levels of the core-layer TFs and bottom-layer TFs are inherently tightly regulated. This suggests that such a tight regulation, along with other regulatory mechanisms such as post-translational modifications or physical interactions with other proteins, might act as a filter to minimize noise propagation down the hierarchy due to any ‘inadvertently’ triggered response. In other words, tight regulation of the core- and bottom-layer TFs via rapid degradation (i.e. shorter protein half-life) would ensure that such TFs are present only in low levels under normal conditions. Their presence in relatively lower levels might facilitate minimization of noise propagation because sufficient levels of TFs may not be present to trigger an appropriate response when transient signals ‘inadvertently’ activate them. Thus the tight regulation of protein levels of the core- and bottom-layer TFs might ensure fidelity and robustness in a regulatory cascade.

Taken together, the findings that we reported suggest that: (i) the higher variability in abundance of top-layer TFs compared with core- and bottom-layer TFs in distinct members of a clonal cell population might permit differential utilization of the same underlying network (Figure 6); and (ii) the tight regulation of core- and bottom-layer TFs might contribute to fidelity in gene expression. Thus the interplay between the dynamics of individual nodes and the topology of the regulatory network would make the underlying network robust and permit at least some members in a population to effectively adapt to (or survive in) changing environments [27].
Figure 6 | A schematic model describing the conceptual framework of differential utilization of the same underlying regulatory network by distinct members of a genetically identical cell population

(A) A model regulatory network showing two regulatory pathways, which will be used to respond to two specific extracellular stimuli. The red, green and blue nodes in the network represent TFs, symbolically representing the inferred top-, core- and bottom-layer TFs in the hierarchical network respectively. (B) Members of a clonal cell population responding to stimulus 1 (top panel). The variability in expression of top-layer TFs (shown as nodes in various shades of red; middle panel) permits differential sampling of the same underlying network by distinct members of a genetically identical population of cells. TFs coloured grey are not expressed at necessary levels, and are shown as inactive nodes. Edges originating from inactive TFs are inactive (shown in grey). A noisy master-regulator TF at the top of the hierarchy would mean that only a subset of a population, in which this TF is expressed at effective levels, will have this TF active. An inactive TF at the top of a hierarchical regulatory cascade will result in the non-expression/inactivation of all downstream TFs and TGs dependent on this TF. Members of a clonal population whose regulatory pathway for a specific extracellular stimulus is active will initiate an effective response when that stimulus is encountered, and those members in whom this regulatory pathway is inactive will be unable to mount an effective response. Although all members in the population are sampling the part of the network necessary to respond to stimulus 1, only a few members (shown as purple and orange cells; bottom panel) are sampling (or poised to sample) the part of the network necessary to respond to stimulus 2. (C) A change in stimulus (from stimulus 1 to 2) causes only those cells that have an active regulatory response pathway for stimulus 2 to effectively respond and survive, whereas the others may mount a late response or do not survive. Alternatively, low expression of top-layer TFs might facilitate cell survival if the pathway regulated by such TFs leads to cell death (e.g. apoptosis). Thus the presence of noisy TFs at the top of the hierarchical regulatory cascade might confer a selective advantage as this permits at least some members in a clonal population to respond to changing conditions. Reprinted by permission from Macmillan Publishers Ltd: Molecular Systems Biology [27] © 2009.
The findings that we reported have implications in synthetic biology experiments aimed at engineering gene-regulatory circuits [111,139,140]. In particular, the dynamics of TFs in terms of their abundance, half-life and noise cannot be ignored because the modulation of these attributes could affect the outcome of a regulatory cascade. The proposed conceptual framework (see Figure 6) from our findings serves as a general model and also has important implications for a number of apparently different, but related phenomena, as outlined below.

**Bacterial persistence or adaptive resistance**

This is a phenomenon where a fraction of a genetically homogeneous microbial population survives upon exposure to stress such as antibiotics [141–143]. Whereas key regulatory proteins, which facilitate random phenotypic switching, have been implicated in this phenomenon [143], we suggest that the altered dynamics and stochastic expression of certain regulatory proteins may alter the susceptibility of an individual by facilitating differential utilization of the underlying network efficiently (Figure 6). This might permit alteration of the phenotype (i.e. the switching in susceptibility states) in an otherwise genetically identical population of cells. This suggestion is supported by studies which have implicated TFs such as PhoU [144] and the two-component signal transduction system DpiAB [145] in this phenomenon.

**Differential cell-fate outcome in response to the same uniform stimulus**

Fractional survival or cell death in clonal cell populations upon drug treatment is a well-known phenomenon in certain diseases such as cancer. Recent studies have shown that this could be achieved by variability in the expression dynamics of key regulatory proteins that determine cell fate such as death or survival [146–150]. In the study by Cohen et al. [146], it was shown that the dynamics of the regulatory proteins, which dictate either cell death or survival, varied widely between individual cancer cells. In line with this, Singh et al. [147] showed that patterns of heterogeneity in the expression level of basal signalling genes in cancer cells can define cell populations that have very different sensitivities to the same drugs. In the other study by Spencer et al. [148], it was shown that naturally occurring differences in the levels or states of proteins regulating apoptosis are the primary causes of cell-to-cell variability in the timing and probability of death in individual members of the population upon induction of apoptosis. We suggest that this dynamic variability in expression level of key regulatory proteins might permit differential sampling (i.e. the survival network or the apoptotic network) of the same underlying regulatory network (governing all cells) by different members in a clonal population (Figure 6). This differential network utilization might result in divergent cell-fate outcomes among different individuals in an otherwise identical cell population subjected to the same uniform stimulus.

**Phenotypic variability in fluctuating environments**

When organisms experience fluctuating environments, individuals of the same population may exhibit very different phenotypes. This may be achieved either by sensing followed by response or through the generation of diversity by random switching between different states [126,151,152]. Although variation in expression levels of certain regulatory proteins has been implicated in this phenomenon, we suggest that such variation in expression levels might allow stochastic switching between phenotypes by poisoning certain individuals or permitting the sampling of relevant parts of the same underlying network (Figure 6). This suggestion is supported by elegant studies. The first one found that, upon fluctuating nutrient starvation, some yeast cells sporulated, while others delayed sporulation [153]. This variability was shown to be governed by the variation in the production rate of the meiotic master regulator Ime1p and its gradual increase over time. In another set of studies, the probabilistic and transient differentiation of *B. subtilis* cells into the state of competence was analysed [154–156]. It was shown that variation in the expression level of the TF ComK determines the frequency of the differentiation events. The increased variability in expression levels of these key regulatory proteins, in turn, might permit different individuals in the same population to activate relevant parts (i.e. the sporulation subnetwork or the competence subnetwork) of the same underlying global transcriptional network to exhibit phenotypic variability (i.e. to sporulate or not; to exhibit a competent or vegetative state).

**Cellular differentiation and development**

During the course of stem-cell differentiation and development, it is well known that maintaining the right balance of combinations of key TFs and appropriate signalling environment in space and time dictates: (i) lineage specification of the progenitor cells (e.g. myeloid lineage commitment from haemopoietic stem cells [157]); and (ii) formation of distinct spatial patterns of cell types during organ development (e.g. cell-fate specification in neural development [158]). In this context, the inherent dynamics of TFs, such as cell-to-cell variation in expression levels, might play an important role in development, stem cell maintenance and differentiation [136,159–163]. Whereas dedicated circuits that filter noise in expression may be required for certain processes to be robust and reproducible [161,164], we suggest that distinct TF dynamics, as dictated by the position in the hierarchy of the transcriptional network, might provide the flexibility for individuals in a population to initiate distinct responses and sample distinct networks that permit lineage commitment when the appropriate signalling environment is experienced (Figure 6). The implications of our findings [27] assume significance particularly in the light of the discovery that transcriptome-wide noise of specific TFs controls lineage choice during stem cell differentiation [165].
In conclusion, investigating the dynamics (e.g. cell-to-cell variability in abundance, half-life of transcripts and proteins) of individual nodes, circuits and the network as a whole together will be critical to obtain a better understanding of how (i) cells adapt to changing environments, (ii) different phenotypic outcomes are mediated in clonal cell populations, and (iii) mutations that disrupt the dynamics of key regulatory proteins may influence disease conditions.

**Outlook and perspectives**

In the present review, I introduced the concept of TRNs and have discussed how representing the transcriptional regulatory system of an organism as a network could provide us with a better understanding of the complexity of gene regulation on a genomic scale. Specifically, I discussed research in the last decade and have highlighted general principles of network structure, evolution and dynamics. In this section, I discuss some of the major challenges and important directions for future research and describe how our understanding of the structure, evolution and dynamics of gene networks are already being exploited in different ways.

**Quantitative modelling of gene networks**

Whereas experimental advances in sequencing are providing us with an avalanche of information about the repertoire of genes and their expression levels across different conditions from diverse microbes and microbial communities, one of the fundamental challenges for the future would be to develop a conceptual and computational framework to integrate all of these data and to quantitatively model how individual genes are regulated within a cell in different contexts such as stress, during infection or in the presence of a particular food source. In this direction, computational and experimental approaches that model regulation of individual genes at high resolution [166,167] or changes in the structure of the entire regulatory network of an organism [101,103] are already being investigated. A key advance would be to investigate different biological systems such as the DNA-damage response or the stress response from diverse organisms, develop new methods for investigating network dynamics and to uncover general principles through comparative analysis.

**DNA structure, modification, packing and gene networks**

Although several studies have modelled gene-regulatory networks involving TF-DNA interactions, there is an increasing need to incorporate the information about nucleosome occupancy, histone variants and chromatin modification with information on TF-DNA interactions to obtain a more complete picture of gene regulation in eukaryotes [168–180]. In the case of prokaryotes, the impact of nucleoid-like proteins and DNA structure such as supercoiling on the regulation of gene expression needs to be considered in a more explicit manner. In this context, several studies are already providing insights into the impact of nucleoid-like proteins on gene-network modelling [109,181–185]. In addition, the impact of DNA modification (e.g. methyl and 5-hydroxymethyl modification [186,187] of cytosine or methyl modification of adenine [188]) and the role of DNA structure such as G-quadruplexes or Z-DNA on transcriptional regulation needs to be investigated systematically. Thus another major direction of future research would be to integrate large datasets describing DNA structure, organization and modification state with the gene-regulatory network to uncover general principles of transcriptional regulation.

**Natural variation of gene networks**

The ability to sequence different strains of the same species or different individuals from the same population is providing us with a wealth of information about natural variation in the genomic sequences of different organisms (e.g. *Mycobacterium leprae* [189], *E. coli* [190,191], yeast [192,193] and humans [194]). Such variation might involve single nucleotide changes [195] or structural alterations such as insertion and deletion of sequences through TEs and HGT [196]. These events affect not only protein coding regions, but also intergenic regions and hence may influence the expression of relevant genes [197,198]. For example, it was shown recently that the gain of a regulatory interaction through mutations in the promoter region of *Salmonella enterica* serotype Typhimurium strains allowed the regulation of a virulence gene [199]. This feature conferred a fitness advantage on those strains and permitted them to adapt better to the host environment [199]. Given the fluid nature of bacterial and eukaryotic genomes, another important future direction would be to understand natural variation in entire gene circuits (not only regulatory interactions) within distinct populations of the same species. Such an understanding can provide fundamental insights into the emergence of pathogens and has implications for human health and disease [66,196].

**Noise and gene networks**

Non-genetic cell-to-cell variation in gene expression (i.e. noise) has been another exciting area that has gained attention recently [129,200]. Such stochastic variation in a cell population can be beneficial where phenotypic diversity is advantageous, but detrimental if homogeneity and fidelity in cellular behaviour is required. Recent work in this direction has shown that different circuits have the potential to either amplify or buffer noise [129,200]. For instance, it was shown recently that whereas seemingly different alternative circuits can provide similar patterns of outputs in gene expression, the impact of fluctuations in protein levels was shown to be an important determinant of why some circuits were selected in evolution [201]. An important challenge in this direction would be to understand the interplay between network structure (at both the local and global levels) and the noise level of individual genes in such networks [27]. Gaining a better understanding of how gene circuits could influence stochasticity in gene expression...
will have a significant impact in understanding how noise could influence phenomena such as incomplete penetrance [202,203], gene-expression divergence [204,205], bacterial persistence or adaptive resistance [142,206], differential cell-fate outcome in response to the same uniform stimulus [154], phenotypic variability in fluctuating environments [126], and finally cellular differentiation and development [155,156].

**Engineering gene circuits**

Another major challenge would be to exploit the knowledge gained about regulatory networks to engineer gene circuits with defined properties (e.g. tuneable and scalable circuits [207–210]) for different applications. In this context, several groups have made important contributions and synthetic gene circuits are already being exploited in medicine (e.g. engineering interactions between bacterial and human cells [211,212] or for human therapeutic applications [213]), bio-energy (e.g. production of fatty-acid-derived fuels [214]), bioremediation (e.g. to harness the concentration gradient of metals [215]), laboratory applications (e.g. creation of bacterial strains resistant to specific antibiotics for selection experiments [216,217]) and in biotechnology (e.g. for the production of proteins [218]). For a more detailed and current account of synthetic biology and engineering of gene circuits, the reader is referred to the following reviews by Chin [219], Kiel et al. [220], Tigges and Fussenegger [221] and Lu et al. [222].

**Genome organization and gene networks**

Advances in microscopy and genome-scale approaches are revealing, for the first time, the way in which genetic material is organized within the nucleus in eukaryotes [223–228]. This new understanding of the three dimensional organization of genetic material raises a fundamental question as to whether the requirements to regulate genes influence where they reside on the chromosome. In this direction, several studies have provided evidence that this is indeed the case. For instance, we and other groups have shown that the genes in *E. coli*, yeast, humans and mouse are organized in a highly ordered manner across and within the linear chromosomes, which might permit precise regulation of genes [99,229–233]. In addition, we and others have shown that where exactly a gene lies on the chromosome can influence its expression level [197,198,234,235] and its noise levels [236,237]. Taken together, these findings open several directions primarily aimed at gaining a better understanding of the influence of genome organization on gene-regulatory networks. Insights in this direction will have implications in applications such as genetic engineering, gene therapy and in understanding diseases that involve chromosomal reorganization such as cancer.

In conclusion, this is truly an exciting time for experimental and computational biologists who aim to understand gene-regulatory networks. Especially with the advances in computing and genomic technologies, I foresee the availability of more extensive and detailed maps of transcriptional regulation and other mechanisms of regulation (e.g. riboswitches, microRNA, RNA-binding proteins, post-translational modifications and small RNAs) in a number of organisms. The availability of such information will fuel research that addresses fundamental questions linking different types of regulation [238–240]. All of these advancements collectively have the potential to transform our understanding of gene regulation in the near future.

**Acknowledgements**

I thank Guilhem Chalancon for help with Figures 2 and 3. I also thank Charles Ravarani, Karthikeyan Sivaraman, Aswin Seshasayee, Sarath Chandra Janga, Muxin Gu, Benjamim Lang, Nadja Zaborsky, A.J. Venkatakrishnan and Guilhem Chalancon for reading this paper. I am indebted and thank all of my colleagues, group members, mentors over the several years (particularly Sarah Teichmann and L. Aravind), friends and my family for their continued support and inspiration.

**Funding**

I acknowledge the Medical Research Council and thank Darwin College, Schlumberger Ltd and the EMBO Young Investigator programme for their generous support.

**References**


© The Authors Journal compilation © 2010 Biochemical Society


Ihmels, J., Bergmann, S., Gerami-Nejad, M., Yani, I., McClellan, M., Berman, J. and Barkai, N. (2005) Rewiring of the yeast transcriptional network through the evolution of motif usage. Science 309, 938-940


